Structure and Dynamics of the N-terminal Loop of PsbQ from Photosystem II of *Spinacia oleracea*

Jaroslava Ristvejová[a], Vladimír Kopecký Jr.[b], Žofie Sovová[a], Mónica Balsera[c], Juan B. Arellano[c], Michael Green[d] and Rüdiger Ettrich[a,*]

[a] *Laboratory of High Performance Computing, Institute of Systems Biology and Ecology of AS CR and Institute of Physical Biology of USB, Zámek 136, 37333 Nové Hrady, Czech Republic*

[b] *Institute of Physics, Faculty of Mathematics and Physics, Charles University, Ke Karlovu 5, 12116 Prague 2, Czech Republic*

[c] *Instituto de Recursos Naturales y Agrobiología (CSIC), Cordel de Marinas 52, 37008 Salamanca, Spain*

[d] *The College of New Jersey, 2000 Pennington Road, Ewing, NJ 08628-0718, U.S.A.*

[*]Corresponding author. Fax: +420 386 361 279; *E-mail address:* ettrich@greentech.cz (R. Ettrich)

Abstract

Infrared and Raman spectroscopy were applied to identify restraints for the structure determination of the 20 amino acid loop between two β-sheets of the N-terminal region of the PsbQ protein of the oxygen evolving complex of photosystem II from *Spinacia oleracea* by restraint-based homology modeling. One of the initial models has shown a stable fold of the loop in a 20 ns molecular dynamics simulation that is in accordance with spectroscopic data. Cleavage of the first 12 amino acids leads to a permanent drift in the root means square deviation of the protein backbone and induces major structural changes.

Keywords: Photosystem II, PsbQ, *Spinacia oleracea*, molecular dynamics, infrared spectroscopy,

*

Abbreviations: FTIR, Fourier-transform infrared; MD, molecular dynamics; PSII, photosystem II

In oxygenic photosynthesis sunlight is converted into chemical energy. This takes place in the thylakoid membrane of green plants, algae and cyanobacteria, in which PSII is the complex that performs light-driven oxidation of water, with reduction of the plastoquinone pool and release of molecular oxygen. PSII is a multi-component protein complex that comprises more than 25 subunits, most of which are embedded in the thylakoid membrane [1]. Water splitting, which gives rise to molecular oxygen, is performed in a cluster of four Mn ions located on the lumenal side of PSII and Ca^{2+} and Cl^{-} ions are required for optimal activity of this water-oxidase complex. The function of Ca^{2+} and Cl^{-} is modulated by the presence of three extrinsic proteins at the lumenal surface [2], which are named PsbO, PsbP and PsbQ in higher plants and create the correct ionic environment during water oxidation. Recently the structure of the recombinant PsbQ protein of PSII from *Spinacia oleracea* was solved by X-ray at a resolution of 1.49 Å [3]. However, despite the high resolution, the loop in the N-terminal region from residue 14 to 33 showed significant disorder in the crystal structure and no structure for this region could be assigned. This natural flexibility, at least when PsbQ is free in solution, leaves MD simulations as the only reasonable alternative to study the behaviour of this loop. In this paper we combine MD analysis with data gained from vibrational spectroscopy to present a structural model for the complete PsbQ protein of PSII from *Spinacia oleracea*.

Materials and methods

Protein preparation. Recombinant PsbQ protein was overexpressed in *Escherichia coli* BL21(DE3)pLysS transformed by JR2592 vector (B96 cells) [4]. The PsbQ was concentrated using Centricon 10 (Millipore) to a concentration of 6 mg/ml. To test the purity of the sample 12% SDS-PAGE electrophoresis was used. Mass of the protein (17 kDa) was determined by nano-ESI-TOF-

MS (Mainer, Applied Biosystems).

FTIR spectroscopy. Infrared spectra were recorded with a Bruker IFS-66/S FTIR spectrometer using a standard source, a KBr beamsplitter and an MCT detector. Generally, 4000 scans were collected with 4 cm^{-1} spectral resolution and a Happ-Genzel apodization function. The prepared samples were measured at room temperature in a CaF_2 BioCell™ (BioTools®) with 10 µm path length. Spectral contribution of the buffer was corrected following the standard algorithm [5].

Raman spectroscopy. Raman spectra were recorded in a standard 90° geometry on a multichannel instrument based on a 600 mm single spectrograph with 1200 grooves/mm grating (Jobin-Yvon), a holographic notch-plus filter (Kaiser Optical Systems) and a liquid nitrogen cooled CCD detection system (Princeton Instruments) having 1024 pixels along dispersion axis. The effective spectral slit width was set to ~4 cm^{-1}. Spectra were averaged from 300 exposures of 120 s each to produce the traces of highest quality. Samples in 10 µl capillary micro-cell were excited with 514.5 nm/100 mW line of an argon ion laser (Coherent Innova 300) and kept in 4 °C during the experiment. The spectra were treated according to [6], subsequently the final spectra were smoothed using 9-point Savitsky-Golay algorithm and normalized to the 1450 cm^{-1} δCH_2 band as an internal standard.

Molecular modeling. The recently solved high resolution X-ray structure of the recombinant PsbQ protein of PSII from *Spinacia oleracea* (pdb: 1VYK, 1.49 Å) was used as a template for homology modeling. The only difference between our structure and 1VYK are the 20 missing residues from 14 to 33 near the N-terminus. Thirty three-dimensional models comprising all non-hydrogen atoms were generated by the MODELLER7 package and refined by a short simulation annealing protocol [7]. The final proposed model was solvated in single point charge water and four chloride counter-ions were added. The system was equilibrated for 250 ps with positional restraints applied on the protein atoms to allow the solvent to relax. The production MD simulations, without any restraints, were 20 ns long and were run with GROMACS version 3.2 [8; 9] using the gmx force field, with a 5 fs time step (dummy hydrogens are used). SETTLE (for water) and LINCS were used to constrain covalent bond lengths, and long-range electrostatic interactions were

computed with the Particle-Mesh Ewald method. The temperature was kept at 300 K by separately coupling the protein and solvent to an external temperature bath (τ = 0.1 ps) [10]. The pressure was kept constant at 1 bar by weak coupling (τ = 1.0 ps) to a pressure bath. The protein proved to be stable during simulation. In a second simulation the starting structure was truncated, the first 12 amino acids were deleted to examine the influence of the β-sheet structure at the N-terminal region on the loop flexibility. The conditions of this second run were chosen exactly the same as with the first. For futher information about molecular modeling in combination with vibrational spectroscopy see ref. [11, 12].

Results and discussion

A three dimensional model of the complete structure of the recombinant PsbQ protein of PSII from *Spinacia oleracea* including all non-hydrogen atoms was generated and refined by long MD simulations in water (20 ns) (Fig. 1). Homology models for further MD simulations were selected by the following criteria: Modeller objective function lower than 800, a *g*-factor above 0, and not more than 1 aminoacid in the forbidden region of the Ramachadran diagram. By that 7 models were selected for MD. Five out of the seven selected models lead to changes in the secondary and tertiary structure of the helix bundle during the MD run that significantly deviated from the crystal structure. Taking into account the high resolution of the crystal structure a correct fold of the N-terminal loop should not change the overall protein structure, which was true for the finally selected structure. One could admit that a crystal might not necessarily represent a solution structure, however in this particular case for PsbQ, Raman and FTIR spectroscopy nicely show that the percentage of the experimentally determined secondary structure in solution matches very well the one of the crystal and so the MD model for PsbQ in solution should not be far from its X ray structure. Both models that did not change the overall secondary structure content significantly lead to the same fold of the loop, but one of them did not show yet equilibrium determined by RMSD during 20 ns, so in fact, MD gave a single model. The stereochemistry of the final proposed model is favourable as indicated by a good quality of stereochemistry and by torsion angles Φ and Ψ

(94.5% residues in the most favourable regions). The overall *g*-factor shows a value of 0.09. The *g*-factor should be above –0.5 and values below –1.0 may need investigation. A constant RMSD value and the stable radius of gyration show that the system reached an equilibrium state (data not shown).Whereas the RMSD in equilibrium oscilates around a value of 1.8 Å from the initial minimized structure for the structure from residue 40 to 149, which in fact is a value that shows a normal dynamic behaviour; the N-terminal loop shows values slightly above 6 Å, indicating the heavy refolding within the first 7 ns of the MD run. The root mean fluctuation in equilibrium state between 10–20 ns shows nicely the larger movement/fluctuation in the loop (Fig. 2). Previously reported estimation of secondary structure content from FTIR and circular dichroism spectroscopy [4] showed a slight discrepancy, therefore a different and more complex type of FTIR spectrum analysis comprising amide I and II bands (Fig. 3) was applied together with measurements of Raman spectrum (Fig. 4) and subsequent analysis in the amide I band region. The excellent agreement of all spectroscopic results with the secondary structure types of the model according to the Kabsch–Sander algorithm is shown in Table 1. Moreover, Raman spectroscopy can provide structural insights of the protein in solution. All of them fit well with the recently X-ray solved structure of PsbQ and one delicate detail can well document quality of the model – the Raman band at 1553 cm^{-1} corresponds to 100° of the torsion angle $|\chi^{2,1}|$ of the single Trp residue [13] which is also in perfect agreement with the model. The final proposed model consists of 4 helices and a parallel β-sheet anchoring the loop near the N-terminus. The helices and the parallel sheet originate from the template structure 1VYK. The loop between Leu13 and Thr34 (Fig. 1) missing in the crystal structure enlarges the β-sheet by adding one small strand formed by two amino acids as calculated by molecular modeling with consequent MD in agreement with secondary structure estimation by vibrational spectroscopy. Within the loop we can find Arg27, a charged residue fully conserved in the PsbQ sequence subfamily of higher plants. Arg27 in the time course of 20 ns has three main interactions with the rest of the protein. H-bond analysis shows that Arg27 has regularly H-bonds with Thr46, Glu47 and Glu21. Hereby we can examine that 2/3 of the occurring H-bonds are within the loop with Glu21, and 1/3 with the helical residues Thr46 and Glu47. As a result of

these weak interactions with the rest of the protein, the loop between Leu13 and Thr34 has a random coil structure with a distinct stable fold; however, its flexibility is much higher than of ordered secondary structures, which seems to be the reason why it cannot be resolved by X-ray diffraction. A key motif in the N-terminal region is the polyproline type II pattern formed by residues 9–12. Together with Pro42, Pro44 and Pro45 the polyproline type II pattern forms a gate for the returning loop that does not show any hydrogen bonding with the prolines and is held in position by steric hindrance only (Fig. 1A). The root mean square fluctuation graph in figure 2 shows a first minimum with the lowest rmsf values at the position of Arg3 and Ile5.This minimum is due to the hydrogen bonding that stabilizes these aminoacids and the surrounding ones, which is true also for the large minimum at Phe38. The second large minimum is at the position of Gln25 and Ala26, which are exactly the amino acids in the gate. The proline residues keep the gate open and prevent both residues sterically from moving to the sides and thus fix them in position. Three out of the four lysyl residues (Lys 90, 96, 101 and 102) which could not be modified with NHS-Biotin when the PsbQ protein is associated with PSII, as reported by [14], are probably oriented to the lumenal facing intrinsic proteins of PSII and are in the same side of the protein as the conserved loop residue Asp24 (Fig. 1B). The probable binding site could be formed by this lysyl rich region of the helix bundle and the N-terminal loop region around Asp24 and thus would contain a large positively charged region and a small negatively one. Lys96 is on the opposite side to the other three lysyl residues and the conserved loop residue Asp24. However, its distance to the loop residues Thr20 and Glu21 varies in the molecular simulation from 6–12 Å and 7–13 Å, respectively (Fig. 5). We hypothesize that after binding to PSII the loop loses its high flexibility and bends in the direction of Lys96 with Thr20 and Glu21 interacting with this residue and so burying it under the accessible surface (Fig. 5). Thus Lys96 could probably behave as a molecular hook holding Glu21 by a salt bridge. In that case it would not be possible to modify the residue with NHS-Biotin when the PsbQ protein is associated with PSII.

As known, a specific prolyl endopeptidase for PsbQ, active only at low ionic strength, cleaves the peptide bond between residues Pro12 and Leu13 in the spinach PsbQ protein [15]. The loss of

affinity by PsbQ for PSII when this peptide fragment is absent has given grounds to propose it as the interacting region of PsbQ. The cleavage site involving residues Pro12–Leu13 is exposed to the solvent and presumably easily accessible to the specific prolyl endopeptidase of PsbQ. On the contrary, the first ten residues at the N terminus are rather occluded in the structure of PsbQ, and because of that they possibly interact weakly with PSII. The MD simulation of the truncated construct shows a steadily growing RMSD, 2.0 Å for C_α during the time frame of 20 ns, indicating a larger motion in the loop region that has not yet reached its equilibrium (Fig. 6). Thus we can state that the cleavage of the Pro12–Leu13 bond leads to a profound instability of the large, flexible and extended loop comprising residues 14–34. The new N-terminal region of degraded PsbQ might behave as a fully free arm. Hence, not only the loss of the peptide fragment comprising the first 12 residues of PsbQ but also the unconstrained structure of the new N-terminal region including Asp24 may produce the non-recognition of PsbQ by PSII.

Acknowledgements

Supports from the Institutional Research Concept of the Academy of Science of the Czech Republic (No. AVOZ60870520) and from the Ministry of Education of the Czech Republic (No. LC 06010, No. MSM0021620835, No. MSM6007665808) and the Grant Agency of the Czech Republic (grant No. 206/03/D082) are gratefully acknowledged. This work was also funded by the Spanish Ministry of Education and Science (Project Ref.: BFU2004-04914-C02-02/BMC).

References

[1] J. Barber, Photosystem II: The engine of life, Quart. Rev. Biophys. 36 (2003) 71–89.

[2] A. Seidler, The extrinsic polypeptides of photosystem II, Biochem. Biophys. Acta 1277 (1996) 35–60.

[3] M. Balsera, J.B. Arellano, J.L. Revuelta, J. De Las Rivas, J.A. Hermoso, The 1.49 Å resolution crystal structure of PsbQ from photosystem II of *Spinacia oleracea* reveals a PPII structure in the N-terminal region, J. Mol. Biol. 350 (2005) 1051–1060.

[4] M. Balsera, J.B. Arellano, J.R. Gutierrez, P. Heredia, J.L. Revuelta, J. De Las Rivas, Structural analysis of the PsbQ protein of photosystem II by fourier transform infrared and circular dichroic spectroscopy and by bioinformatic methods, Biochemistry 42 (2003) 1000–1007.

[5] F. Dousseau, M. Therrien, M. Pézolet, On the spectral subtraction of water from FT-IR spectra of aqueous solutions of proteins, Appl. Spectrosc. 43 (1989) 538–542.

[6] R.W. Williams, The secondary structure analysis using Raman amide I and amide III spectra, Methods Enzymol. 130 (1986) 311–331.

[7] A. Sali, J.H. Overington, Derivation of rules for comparative protein modeling from a database of protein structure alignments, Protein Sci. 3 (1994) 1582–1596.

[8] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56.

[9] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: A package for molecular simulation and trajectory analysis, J. Mol. Modell. 7 (2001) 306–317.

[10] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular-dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.

[11] V. Kopecký Jr., R. Ettrich, K. Hofbauerová, V. Baumruk, Structure of human α_1-acid glycoprotein and its high-affinity binding site, Biochem. Biophys. Res. Commun. (2003) 41–46.

[12] V. Kopecký Jr., R. Ettrich, K. Hofbauerová, V. Baumruk, Vibrational spectroscopy and computer modeling of proteins: Solving structure of α_1-acid glycoprotein, Spectrosc.–Int. J. 18 (2004) 323–330.

[13] T. Miura, G.J. Thomas Jr., Raman spectroscopy of proteins and their assemblies, in: B.B. Biswas, S. Roy (Eds.), Subcellular Biochemistry – Proteins: Structure, Function, and Engineering, Vol. 24, Plenum Press, New York, 1995, pp. 55–99.

[14] G.D. Meades, A. McLachlan, L. Sallans, P.A. Limbach, L.K. Frankel, T.M. Bricker, Association of the 17-kDa Extrinsic Protein with Photosystem II in Higher Plants, Biochemistry 44 (2005) 15216–15221.

[15] T. Kuwabara, Characterization of a propyl endopeptidase from spinach thylakoids, FEBS Lett. 300 (1992) 127–130.

[16] F. Dousseau, M. Pézolet, Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands, Biochemistry 29 (1990) 8771–8779.

Tables

Table 1. Secondary structure content (in percentage) of PsbQ estimated from amide I and II bands of FTIR spectroscopy by least square analysis with reference set of 19 protein spectra according to ref. [16], amide I band of Raman spectroscopy by least square analysis with reference set of 15 protein spectra according to ref. [6] and MD modeling. Given standard deviations were calculated by one-out tests of the reference sets. Additionally, our results are compared with previous results from circular dichroism [4].

Figure legends

Fig. 1. PsbQ after 20 ns of MD at 300 K. (A) Four Pro residues (residues 9–12) fulfilling the spatial features of the secondary structure conformation denoted as polyproline type II in connection with Pro42, Pro44 and Pro45 form an open gate for residues Asp24 and Ala25 without directly interacting with these residues by hydrogen bonding. (B) The conserved loop residue Asp24 shows a location on the same side of the protein as three out of the four lysyl residues which are probably orientated to the lumenal facing intrinsic proteins of PSII.

Fig. 2. Root mean square fluctuation of PsbQ during the last 10 ns of MD simulation at 300 K.

Fig. 3. FTIR spectrum of PsbQ in the region of amide I and amide II bands. The dash-dot line represents 2nd derivative (smoothed by Savitski-Golay function at 9 points, i.e. *ca.* 15 cm^{-1}) of the spectrum

Fig. 4. Raman spectrum of PsbQ. The assignments of the bands were provided according to ref. [12], whereas ν corresponds to stretching and δ to bending vibrations.

Fig. 5. Lys96, one of the four lysyl residues which are probably orientated to the lumenal facing intrinsic proteins of PSII, lies on the opposite side as are the other three lysyl residues and the conserved loop residue Asp24, in a distance from 6–12 Å and 7–13 Å to the loop residues Thr20 and Glu21, respectively.

Fig. 6. RMSD of C_α of the truncated construct during the 20 ns MD run at 300 K. The RMSD value is continuously drifting, indicating irreversible conformational changes due to the cleavage.

Figure1

Click here to download high resolution image

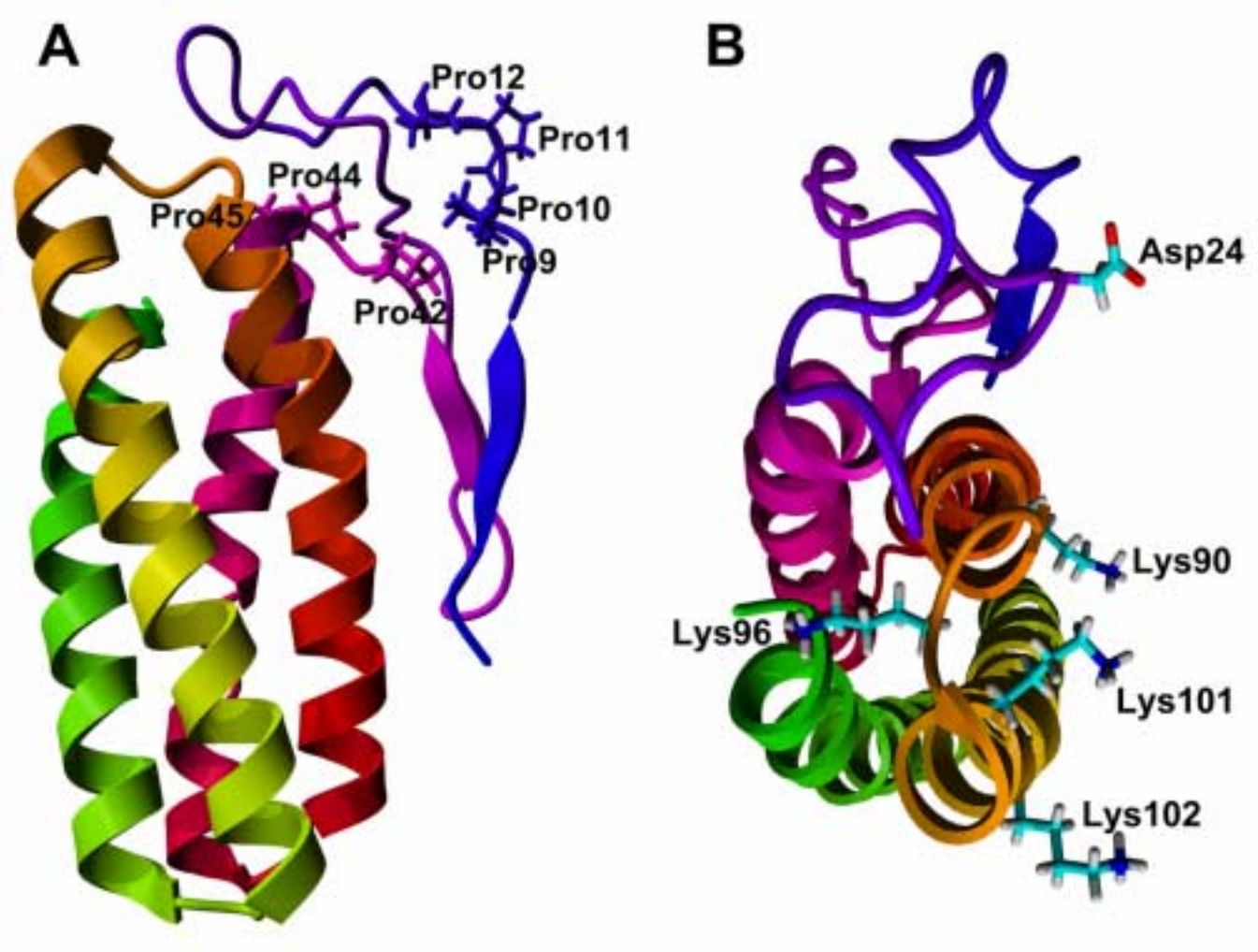

Figure2

Click here to download high resolution image

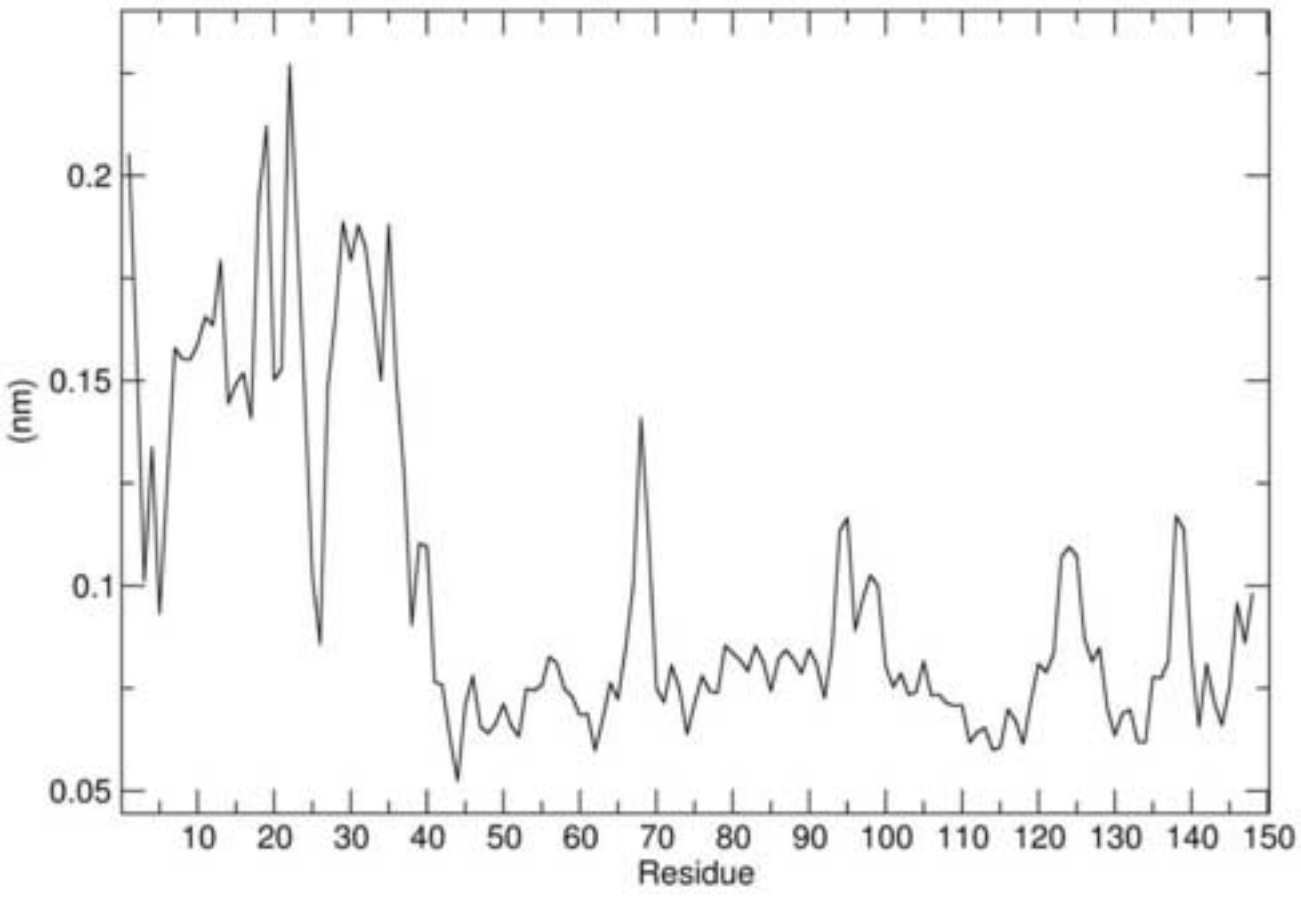

Figure3

Click here to download high resolution image

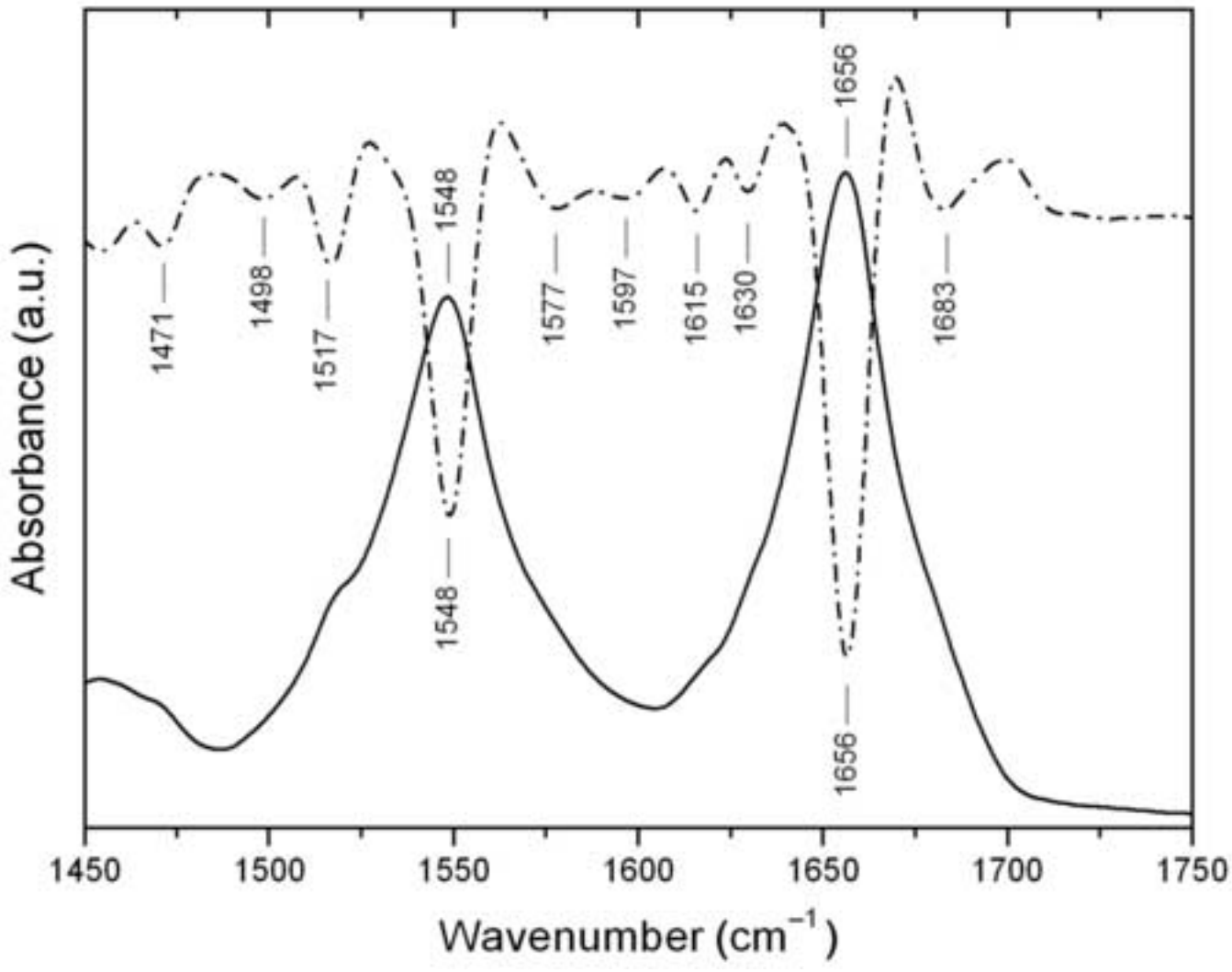

Figure4

Click here to download high resolution image

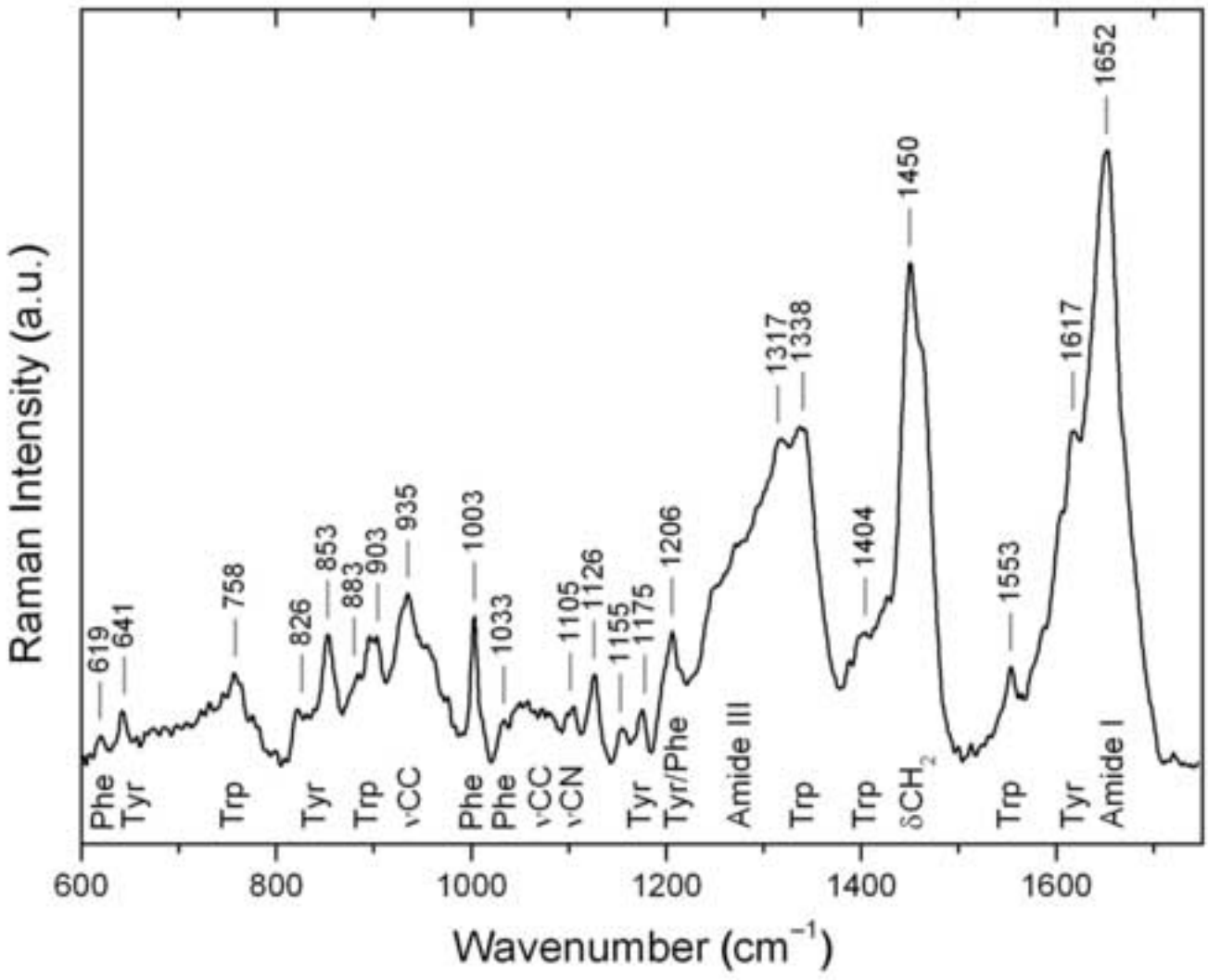

Figure5

Click here to download high resolution image

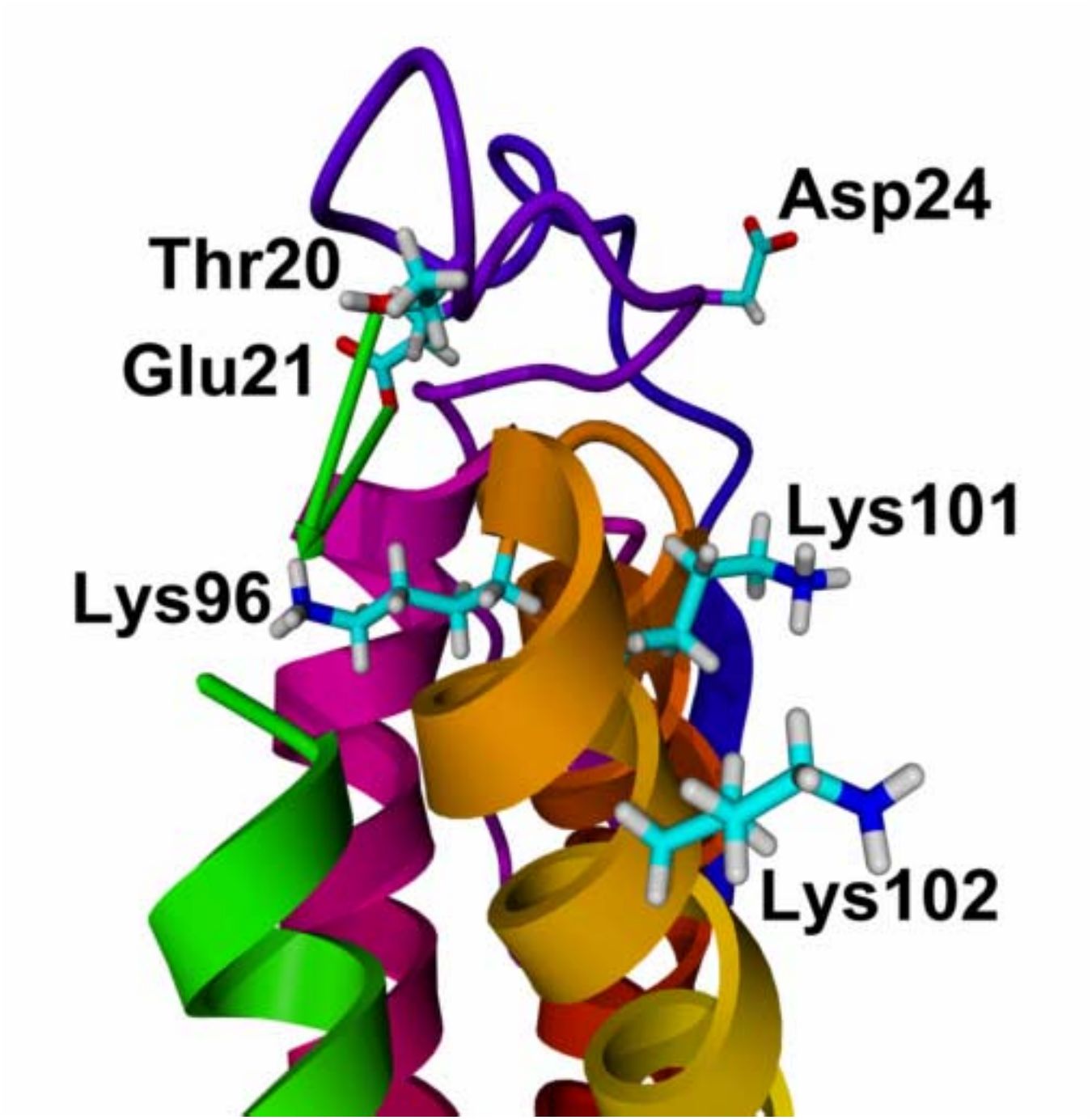

Figure6

Click here to download high resolution image

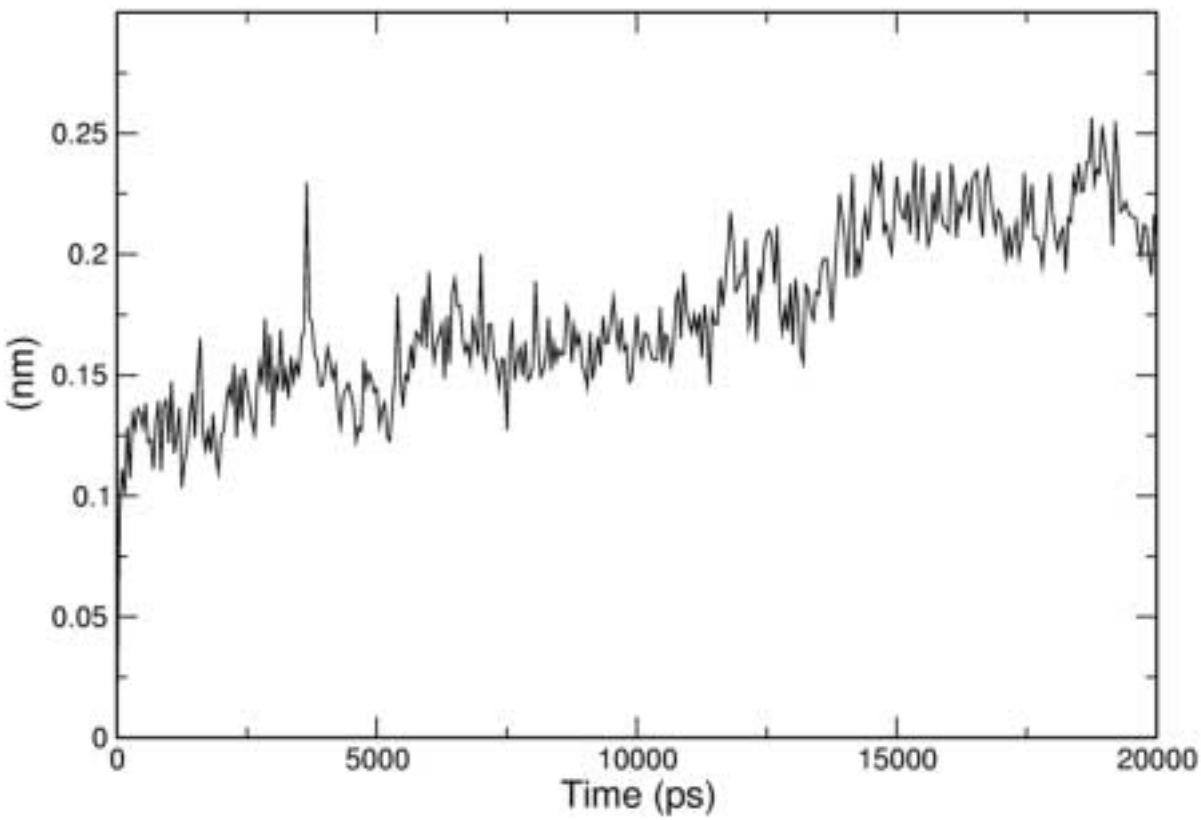

Table1

Structure type	FTIR	Raman	Circular Dichroism [4]	MD Modeling
α-helix	63 ± 10	65 ± 4	64 ± 9	63
β-sheet	5 ± 8	4 ± 4	7 ± 4	7
β-turn	12 ± 4	8 ± 2	—	7
bend	6 ± 4	—	—	5
other	16 ± 6	19 ± 2	29 ± 7	18